Abstracts

Phase Vocoder
J. L. Flanagan and R. M. Golden

A vocoder technique is described in which speech signals are represented by their short-time phase and amplitude spectra. A complete transmission system utilizing this approach is simulated on a digital computer. The encoding method leads to an economy in transmission bandwidth and to a means for time compression and expansion of speech signals.

The voice without soul: origin and history of the Vocoder
Paolo Zavagna

A review of the cultural origins of the speaking machines is proposed. These origins are identified in the separation between soul and body, proposed by Descartes. The speaking machines, and then the vocoder, are substantially formed by a controlled part and an automatic part; in the voder, the synthetic speaker is directly controlled by a human being, conversely in the vocoder an authomatic process is implemented. Through historical examples such as the 'automata' by Mical, Kratzenstein, von Kempelen, Wheatstone, Helmoltz, until the invention of the Dudley's vocoder, are described a growth and a stratification of devices and synthesizers applied to varied fields (signal representation, encrypted transmission, medical apparatus, cybernetic, music). The historical period treated ideally ends with Dolson's article written in 1986, "The Phase Vocoder: A Tutorial". Through musical examples is described the origin of control data handling in composition. The musical application of the phase vocoder describes the distinction between control and synthesis better than other technologies. Both control (parameter values) and synthesis (instruments design) are compositional problems that involve all the electroacoustic music composers.

From Sound Morphing to the Synthesis of Starlight.
Musical experiences with the Phase Vocoder over 25 years
Trevor Wishart

The article reports the author's experiences with the phase vocoder. Starting from the first attempts during the years 1973-77 – in connection with a speculative project to morph the sounds of a speaking voice into sounds from the natural world, project subsequently developed at Ircam in Paris between 1979 and 1986 – up to the most recent experiences in 2011-12 associated with the realization of *Supernova*, an 8-channel sound-surround piece, where the phase vocoder data format is used as a synthesis tool.

Phase vocoder and beyond
Marco Liuni and Axel Röbel

For a broad range of sound transformations, quality is measured according to the common expectation about the result: if a male's voice has to be changed in a female's one, there exists a common reference for the perceptive evaluation of the result; the same holds if an instrumental sound has to be made longer, or shorter. Following the argument in Röbel, "Between Physics and Perception: Signal Models for High Level Audio Processing", a fundamental requirement for these transformation algorithms is their need of signal models that are strongly linked to perceptually relevant physical properties of the sound source. This paper is a short survey about the *phase vocoder* technique, together with its extensions and improvements relying on appropriate sound models, which have led to high level audio processing algorithms.

Arbitrary Phase Vocoders by means of Warping
Gianpaolo Evangelista, Monika Dörfler and Ewa Matusiak

The Phase Vocoder plays a central role in sound analysis and synthesis, allowing us to represent a sound signal in both time and frequency, similar to a music score – but possibly at much finer time and frequency scales – describing the evolution of sound events. According to the uncertainty principle, time and frequency are not independent variables so that any time-frequency representation is the result of a compromise between time and frequency resolutions, the product of which cannot be smaller than a given constant. Therefore, finer frequency resolution can only be achieved with coarser time resolution and, similarly, finer time resolution results in coarser frequency resolution.

While most of the conventional methods for time-frequency representations are based on uniform time and uniform frequency resolutions, perception and physical characteristics of sound signals suggest the need for nonuniform analysis and synthesis. As the results of psycho-acoustic research show, human hearing is naturally organized in nonuniform frequency bands. On the physical side, the sounds of percussive instruments as well as piano in the low register, show partials whose frequencies are not uniformly spaced, as opposed to the uniformly spaced partial frequencies found in harmonic sounds. Moreover, the different characteristics of sound signals at the onset transients with respect to stationary segments suggest the need for nonuniform time resolution. In the effort to exploit the time-frequency resolution compromise at its best, a tight time-frequency suit should be tailored to snuggly fit the sound body.

In this paper we overview flexible design methods for phase vocoders with nonuniform resolutions. The methods are based on remapping the time or the frequency axis, or both, by employing suitable functions acting as warping maps, which locally change the characteristics of the time-frequency plane. As a result, the sliding windows may have time dependent duration and/or frequency dependent bandwidth. As an example, in a constant Q frequency band allocation, the ratios of center band frequencies over bandwidth remains constant, so that the frequency bands become wider and wider as center frequency increases, similarly to the frequency distance of 12-tone scale notes or of octaves.

While time-frequency allocation can be performed in an arbitrary way, the ability to reconstruct the original signal from Vocoder analysis data is essential in sound processing and transformation applications. Moreover, even the analysis or the production of spectrograms benefits from the perfect reconstruction property if one needs to be confident that no important information is hidden, which serves to completely describe the signal.